

УДК 004.9

ИНФОРМАЦИОННО-ПОИСКОВЫЕ СИСТЕМЫ В ОБРАЗОВАНИИ

И.Г. Чернов¹, А.А. Агафонова²

¹студент факультета «Математики и информационных технологий», группа АИС-41, Стерлитамакский филиал «Башкирский государственный университет», г. Стерлитамак, Россия, e-mail: andakm@rambler.ru

²студент факультета «Математики и информационных технологий», группа ФМ-31, Стерлитамакский филиал «Башкирский государственный университет», г. Стерлитамак, Россия, e-mail: andakm@rambler.ru

Аннотация. В работе рассматриваются основные принципы построения информационно-поисковых систем (ИПС) и их практического применения в образовательном процессе. Рассматриваются различные методики обучения пользователей технологии поиска информации, так как уровень компетентности в области информационного поиска у школьников и студентов остается низким. Приводятся понятия и определения, обсуждаются информационные потребности пользователей. Рассматривается представление информации в ИПС, принципы анализа текстов и индексирования документов, типичные модели и алгоритмы поиска информации. Приводятся основные сведения о классификации документов. Обсуждаются современные словарные, классификационные и метапоисковые ИПС, их практическое применение и критерии эффективности.

Ключевые слова: *информационный поиск, информационно-поисковые системы, дифференциальные признаки, информационные запросы*

INFORMATION RETRIEVAL SYSTEMS IN EDUCATION

Igor Chernov¹, Alena Agafonova²

¹student of the Faculty "Mathematics and Information Technologies", group of AIS-41, Sterlitamak Branch "Bashkir State University", Sterlitamak, Russia, e-mail: andakm@rambler.ru

²student of the Faculty "Mathematics and Information Technologies", group of FM-31, Sterlitamak Branch "Bashkir State University", Sterlitamak, Russia, e-mail: andakm@rambler.ru

Abstract. The paper deals with the basic principles of information retrieval systems (IPS) and their practical application in the educational process. Various methods of teaching users search information technology, as the level of competence in the field of information retrieval in pupils and students remains low. Are concepts and definitions, discusses the information needs of users. We consider the provision of information to the IRS, the principles of text analysis and indexing of documents, the typical model and retrieval algorithms. The basic data on the classification of documents. We discuss modern vocabulary, classification and metasearch IRS, their practical application and effectiveness criteria.

Keywords: *information search, information retrieval systems, distinctive features, information requests.*

Введение. С каждым годом объемы сети Интернет увеличиваются в разы, поэтому вероятность найти необходимую информацию резко возрастает. Интернет объединяет миллионы компьютеров, множество разных сетей, число пользователей увеличивается на 15-80% ежегодно. И, тем не менее, все чаще при обращении к сети Интернет основной проблемой оказывается не отсутствие искомой информации, а возможность ее найти. Как правило, обычный человек в силу разных обстоятельств не может или не хочет тратить на поиск нужного ему ответа больше 15-20 минут. Поэтому особенно актуально правильно и грамотно научиться, где и как искать, чтобы получать желаемые ответы. Чтобы найти нужную информацию, необходимо найти её адрес. Для этого существуют специализированные поисковые сервера (роботы индексов (поисковые системы), тематические Интернет-каталоги, системы мета-поиска, службы поиска людей и т.д.).

Цель работы. В работе рассмотрим основные технологии поиска информации в Интернете, общие черты поисковых инструментов, структуры поисковых запросов для наиболее популярных русскоязычных и англоязычных поисковых систем. Также рассмотрим методику грамотно и эффективно строить поисковые запросы информации.

Материал и результаты исследований. ИПС (информационно-поисковая система) – это система, обеспечивающая поиск и отбор необходимых данных в специальной базе с описаниями источников информации (индексе) на основе информационно-поискового языка и соответствующих правил поиска. Главной задачей любой ИПС является поиск информации, релевантной информационным потребностям пользователя. Очень важно в результате проведенного поиска ничего не потерять, то есть найти все документы, относящиеся к запросу, и не найти ничего лишнего. Поэтому вводится качественная характеристика процедуры поиска – релевантность – это соответствие результатов поиска сформулированному запросу. Знания об информационных ресурсах, видах информации и классификации документов необходимы при формулировании и анализе информационного запроса – ведь его автор должен понимать, какой сегмент информационных ресурсов он хотел бы получить в результате поиска информации. Пользователям необходимо знать, что информация делится на фактографическую и документальную, а документы по степени свертывания – на первичные и вторичные; по целевому назначению – на официальные, научные, учебные, рекламные и др.; по знаковой природе – на текстовые, графические, нотные и др.

Мировые информационные ресурсы разделяются на три сектора:

- деловой информации;
- научно-технической и специальной информации;

• массовой, потребительской информации.

Основными показателями ИПС для WWW являются пространственный масштаб и специализация. По пространственному масштабу ИПС можно разделить на локальные, глобальные, региональные и специализированные.

Локальные поисковые системы могут быть разработаны для быстрого поиска страниц в масштабе отдельного сервера.

Региональные ИПС описывают информационные ресурсы определенного региона, например, русскоязычные страницы в Интернете.

Глобальные поисковые системы в отличие от локальных стремятся по возможности наиболее полно описать ресурсы всего информационного пространства сети Интернет. Их можно разделить на

1) *Уникальные ИПС Yandex, Google, Rambler*

2) *Мета-системы Metabot.ru, Nigma*

3) *Специализированные ИПС.* Например поиск изображений, поиск авиабилетов Skyscanner.ru, поиск отелей Booking.com

Web-технология World Wide Web (WWW) считается специальной технологией подготовки и размещения документов в сети Интернет. В состав WWW входят web-страницы, электронные библиотеки, каталоги, виртуальные музеи и т.д. Каким образом можно сориентироваться в столь масштабном информационном пространстве?» В решении данной проблемы используются поисковые инструменты. Поисковые инструменты-это особое программное обеспечение, основная цель которого – обеспечить наиболее оптимальный и качественный поиск информации для пользователей Интернета. Поисковые инструменты размещаются на специальных веб-серверах, каждый из которых выполняет определенную функцию:

1. Анализ веб-страниц и занесение результатов анализа на тот или иной уровень базы данных поискового сервера.

2. Поиск информации по запросу пользователя.

3. Обеспечение удобного интерфейса для поиска информации и просмотра результата поиска пользователем.

Интерфейс поискового инструмента - страница с гиперссылками, строкой подачи запроса (строкой поиска) и инструментами активизации запроса. Индекс поисковой системы-это информационная база, содержащая результат анализа Web-страниц, составленная по определенным правилам. Запрос-это ключевое слово или фраза, которую вводит пользователь в строку поиска. Для формирования различных запросов используются специальные символы ("", , ~), математические символы (*, +, ?). Схема поиска информации в сети Интернет проста. Пользователь набирает ключевую фразу и активизирует поиск, тем самым получает подборку документов по

сформулированному (заданному) запросу. Этот список документов ранжируется по определенным критериям так, чтобы вверху списка оказались те документы, которые наиболее соответствуют запросу пользователя. Каждый из поисковых инструментов использует различные критерии ранжирования документов, как при анализе результатов поиска, так и при формировании индекса (наполнении индексной базы данных web-страниц). Таким образом, если указать в строке поиска для каждого поискового инструмента одинаковой конструкции запрос, можно получить различные результаты поиска. Для пользователя имеет большое значение, какие документы окажутся в первых двух-трех десятках документов по результатам поиска и насколько эти документы соответствуют ожиданиям пользователя. Большинство поисковых инструментов предлагают два способа поиска – simple search (простой поиск) и advanced search (расширенный поиск) с использованием специальной формы запроса и без нее. Освоение критериев уточнения запроса и приемов расширенного поиска, позволяет увеличивать эффективность поиска и достаточно быстро найти необходимую информацию. Прежде всего, увеличить эффективность поиска можно за счет использования в запросах логических операторов (операций) Or, And, Near, Not, математических и специальных символов.

Так в Yandex можно формировать следующие запросы:

- *ключевое слово **site:адрес сайта*** ищет на данном сайте ключевое слово
- *ключевое слово **site:ru*** будут найдены документы с указанными словами расположенными только в доменной части ru
- ***info:адрес сайта*** позволяет увидеть информацию, известную об указанном сайте.
- ***link:адрес сайта*** показывает все страницы, в которых есть ссылки на указанный сайт
- ***allintitle:ключевые слова*** показывает сообщения, в заголовках которых есть *ключевые слова*. Ключевые слова указываются через пробел.
- ***intitle:ключевые слова*** только первое слово после двоеточия будет в заголовке сообщения, остальные слова в теле самого сообщения.
- ***allinurl:слова*** все слова запроса содержатся только в адресе страницы (url).
- ***inurl:слово1 слово2*** слово1 содержится в адресе страницы, остальные слова – на самой странице.
- ***related:адрес страницы*** описывает страницы, похожие на страницу, адрес которой указан.
- ***define:слово*** выбирает страницы, где есть определения данного слова или словосочетания.

• **Слово число1..число2** выбирает страницы с заданным словом и числом, входящим в заданный интервал.

Рассмотрим поисковую систему Google. Он предлагает множество различных функций, которые могут вам помочь в поиске. Подробную информацию об этих функциях вы можете получить на <http://www.google.ru/intl/ru/help/features.html>.

Как известно, информационный запрос – это словесное выражение информационной потребности. Для достижения успешного результата в процессе поиска информации по конкретному информационному запросу пользователю целесообразно пройти несколько последовательных этапов (они описаны в [1], [2]).

1. Формулировка информационного запроса.
2. Анализ информационного запроса: определение объекта и аспекта поиска с использованием словаря значений дифференциальных признаков.
3. Деление запроса на подзапросы (при наличии нескольких тем).
4. Составление поискового образа запроса (ПОЗ) на языке ключевых слов с использованием логических операций конъюнкции, дизъюнкции и отрицания.
5. Соотнесение значений признаков запроса с дифференциальными признаками ИПС.
6. Отбор круга ИПС, в которых целесообразно искать информацию по заданным критериям.
7. Построение общей стратегии поиска.
8. Составление поисковых формул на языке запросов конкретных ИПС.
9. Переформулирование запроса с использованием синонимичных терминов, ассоциативных, видовых (а нередко и родовых) понятий.

В каждом запросе содержатся информация о том, что нужно найти пользователю (объект поиска), и *те признаки*, по которым следует отбирать информацию (аспект поиска, или критерии отбора).

В общем случае, можно выделить следующие поисковые инструменты для WWW: каталоги, поисковые системы, метапоисковые системы. Каталог поисковая система с классифицированным по темам списком аннотаций со ссылками на Web-ресурсы (классификация, как правило, проводится людьми). Поиск в каталоге очень удобен и проводится посредством последовательного уточнения тем. Тем не менее, каталоги поддерживают возможность быстрого поиска определенной категории или страницы по ключевым словам с помощью локальной поисковой машины. База данных ссылок (индекс) каталога обычно имеет ограниченный объем, заполняется вручную персоналом каталога. Некоторые каталоги используют автоматическое обновление индекса. Результат поиска в каталоге представляется в

виде списка, состоящего из краткого описания (аннотации) документов с гипертекстовой ссылкой на первоисточник. Поисковая машина-поисковая система с формируемой роботом базой данных, содержащей информацию об информационных ресурсах. Отличительной чертой поисковых машин является тот факт, что база данных, содержащая информацию об Web-страницах, статьях Usenet и т.д., формируется программой-роботом. Поиск в такой системе проводится по запросу, составленному пользователем, состоящему из набора ключевых слов или фразы, заключенной в кавычки. Индекс формируется и поддерживается в актуальном состоянии роботами - индексировщиками. В описании документа чаще всего содержится несколько первых предложений или выдержки из текста документа с выделением ключевых слов. Как правило, указана дата обновления (проверки) документа, его размер в килобайтах, некоторые системы определяют язык документа и его кодировку (для русскоязычных документов). Если название и описание документа соответствует требованиям пользователя, можно перейти к его первоисточнику по ссылке. Это удобнее делать в новом окне, чтобы иметь возможность далее анализировать результаты выдачи. Многие поисковые системы позволяют проводить поиск в найденных документах, причем можно уточнить запрос введением дополнительных терминов. Если интеллектуальность системы высока, может быть предложена услуга поиска похожих документов. Некоторые поисковики позволяют провести пересортировку результатов. К наиболее известным машинам веб-поиска относятся Google, Yahoo, Alta Vista, Excite, Hot Bot, Lycos. Среди русскоязычных можно выделить Yandex, Rambler. Метапоисковые системы (поисковые службы) – системы, способные послать запросы пользователя одновременно нескольким поисковым серверам, затем объединить полученные результаты и представить их пользователю в виде документа со ссылками. Наиболее популярная в мире система мета-поиска Search.com. Объединенный поисковый сервер Search.com компании CNET, Inc. включает в себя почти два десятка поисковых систем. Отрицательной стороной метапоисковых систем можно назвать их нестабильность. Кроме основных поисковых инструментов можно представить еще несколько: Подборки ссылок – это отсортированные по темам ссылки. Они достаточно сильно отличаются друг от друга по наполнению, поэтому чтобы найти подборку, наиболее полно отвечающую интересам пользователя, необходимо ходить по ним самостоятельно, чтобы составить собственное мнение. Преимуществом такого вида поисковых инструментов является их целенаправленность. Обычно подборка включает в себя редкие интернет-ресурсы, подобранные конкретным Web-мастером или хозяином интернет-странички. Базы данных адресов–

это специальные поисковые сервера, которые обычно используют классификации по роду деятельности, по выпускаемой продукции и оказываемым услугам, по географическому признаку. Иногда они дополнены поиском по алфавиту. В записях базы данных хранится информация о сайтах, которые предоставляют информацию об электронном адресе, организации и почтовом адресе за определенную плату. Система поиска FTP-файлов – это особый тип средств поиска в Internet, который позволяет находить файлы, доступные на «анонимных» FTP-серверах. Протокол FTP предназначен для передачи по сети файлов, и в этом смысле он функционально является своеобразным аналогом Gopher. Основным критерием поиска является название файла, задаваемое разными способами (точное соответствие, подстрока, регулярное выражение и т.д.). Данный тип поиска не может соперничать по возможностям с поисковыми машинами, так как содержимое файлов никак не учитывается при поиске, а файлам можно давать произвольные имена. Тем не менее, если требуется найти какую-нибудь известную программу или описание стандарта, то с большой долей вероятности файл, его содержащий, будет иметь соответствующее имя, и можно найти его при помощи одного из серверов FTP Search: FileSearch ищет файлы на FTP-серверах по именам самих файлов и каталогов. Если пользователь ищет какую-либо программу или еще что-то, то на WWW-серверах он находит их описание, а с FTP-серверов можно перекачать их к себе.

Вывод. Результатом изучения данной темы является формирование сложного умения – эффективного поиска информации, включающего комплекс простых умений: формулировать и анализировать запрос, сравнивать ИПС, «переводить» запрос на язык конкретной ИПС, выстраивать стратегию поиска. Эти умения должны основываться на базовых знаниях, которые студенты приобретают в процессе изучения разных тем курса. Требуется последовательное развертывание предметного содержания, выполнение цикла разнообразных лабораторных и внеаудиторных заданий.

Работа выполнена под научным руководством к.ф.-м.н., доц. Акимова А.А.

ЛИТЕРАТУРА

1. Витухновская А.А., Ключенко Т.И. Формализация информационного поиска в библиотеке/А.А. Витухновская // Научн. и техн. б-ки СССР. –1991. –№ 10. –С. 27–35.
2. Витухновская А.А. Обучение студентов педвузов технологии решения информационно-поисковых задач/ А.А. Витухновская // Информатика и образование. – 2003. –№ 7. – С. 96–101.
3. Акимов А.А. Электронное учебное пособие «Информационные технологии в решении экономических задач»/ А.А.Акимов, Г.Р. Галиаскарова, Р.Г. Идрисов // Хроники объединенного фонда электронных ресурсов Наука и образование. – 2014. – №10(65). – С. 30.